AFTERNOON SESSION

(1:03 p.m.)

DR. LEE: We have a number of guests at the table. I think that we know both of them, but for the record, would you please introduce yourselves?

DR. ENDRENYI: Laszlo Endrenyi, University of Toronto.

DR. YACOBI: I'm Avi Yacobi from Taro Pharmaceuticals.

DR. LEE: And we have Professor Benet, allegedly in his office.

DR. BENET: I am here.

DR. LEE: Okay. Thank you, Les. Les said that he could see us but we could not see him.

The agenda for this afternoon's session is on individual bioequivalence, and we have plans for about 90 minutes on background information. We'll go for a break, and the committee is going to deliberate on the issues at 2:45. And I'd like to draw the attention of the committee to the five topics. Marv Meyer will be leading the discussion and he's going to tell us exactly what to do.

(Laughter.)

DR. LEE: Dr. Lesko, are you ready?

DR. LESKO: Yes.

DR. LEE: Please.

DR. LESKO: Good afternoon, everybody. The purpose of my being up here at the moment is to introduce the topics for this afternoon. I'll provide a little bit of a background to the discussion topics and some of the rationale for bringing these topics to the committee.

Average bioequivalence represents the current and traditional standard for the approval of generic drug products and products post-marketing after some changes in their manufacturing.

It's been used by the FDA to analyze clinical trials for the marketing of thousands of generic drugs. The agency recognizes that in some cases there is a need for other standards or alternative standards and for a few drugs, such as those defined by class I of our biopharmaceutic classification system, in vivo studies are waived and market access is granted on the basis of in vitro studies.

There's a large amount of empirical evidence that suggests that generic drugs are used regularly without serious problems of safety and efficacy, and the agency feels confident in the therapeutic equivalence of these products.

Individual bioequivalence represents an improved standard in the agency's mind, and it was proposed by FDA as an improvement on the study design, the

informativeness, and the method of analysis of BE studies. You heard a little bit this morning about the differences between average and individual bioequivalence. This approach takes into account within-subject variability for both the test and reference product. It detects signals that may represent a subject-by-formulation interaction, and it allows for scaling of the bioequivalence limits.

It's been a controversial topic with many debates and public discussions, to say the least. Through these public discussions and debates we resolve many of the issues associated with this approach, but as of today it has not been universally accepted in the scientific community, or by other regulatory agencies.

A less thorough discussion of this topic in front of the group was back in 1999. At that time the recommendation of the committee was that they had concerns with the new criterion, and recommended use of the ABE criterion for market access unless there is a compelling reason not to. This was reflected, I believe, because of some insufficient data at the time to replace the old standard of average bioequivalence with the new one, where there may be some risks that were either unknown or unappreciated at the time.

We subsequently about a year ago came out with a general BA/BE guidance, and the focus of this afternoon

will be on one section of that guidance. The section that is in focus is the one that deals with the comparison of BA measures in BE studies. It's section IV of the guidance. And the key words in that section are the ones I've italicized and bolded. It says, however, sponsors have the option to explain why they would use another criterion other than ABE. One of the examples might be highly variable drugs and the use of replicate design studies.

However, what this language allows for is an opening, in a sense, for using individual bioequivalence for allowing market access of a generic drug.

A few sponsors have actually requested a priori in their bioequivalence study protocols that the agency use IBE to allow scaling and to allow access to the marketplace.

So, the agency has a dilemma in a sense in making the decision on market access based on the scientific evidence presented by these replicate design studies, and we'd like to bring some of this data to the committee for their evaluation today.

That leads me to the first discussion topic, and it has to do with, is it reasonable and appropriate for FDA to use ABE for market access unless there is compelling reason not to during an interim period for another year from today until we make the final decision to use IBE for

market access.

We're sort of one year post the guidance, and we've acquired about 20 replicate data sets since that time in ANDAs and NDAs, which you'll hear about. We'll be presenting that data to the committee today, and from your look at that data, whether that data provides any new insights into the use of IBE. We feel that this discussion topic in a sense confirms the current situation and doesn't necessarily represent anything new.

The reason we prefer to stay at our status quo is that we still have some concerns about using IBE for market access and some unintended consequences perhaps of this criterion. Of the data sets we've accrued in the past year, most of them pass both average bioequivalence and individual bioequivalence, and as a result are not very informative. We focused on those data sets where one of the criteria passes and one fails. That's where we want to try to discern the differences in the behavior of these criterion.

One of the things that presents a dilemma for us is a situation when ABE fails our current standard but IBE passes, and we have this example in two cases. In the NDA data that we've sent you, drug number 6 represented this phenomenon, and in the ANDA data set, drug number 2 represented this phenomenon.

When we have this situation, it raises some questions. It raises questions, for example, in the ANDA drug number 2. Is this product switchable, and does IBE assure that? In this example the mean test-to-reference ratio was 88.5. We estimate that up to a 15 percent difference in the test-to-reference ratio can pass ABE, so there's nothing remarkable there. This drug in fact may have passed average, had it been powered with more subjects.

The within-subject variability was pretty much similar. The test had a modestly higher variability, but the subject-by-formulation interaction was what we considered important in our guidance, when the value of that SxF exceeds 0.15.

So, some of the concern with the criterion is that it's designed to identify signals of a subject-by-formulation interaction. Unless we have some other evidence to the contrary in this study, one might assume that this is a real signal of a subject-by-formulation interaction. Yet, in the face of that, while we succeeded in detecting it, the IBE criterion says to pass the product.

Furthermore, like most of the studies that are in the new data we sent you, the subject population has been healthy. All male volunteers. We do recommend in the

guidance a heterogeneous population, and as a result we feel that the all-male volunteer population may tend to reduce the frequency of the subject-by-formulation interaction.

The guidance states that the mean test-to-reference ratio should fall between 80 to 125, and in this example there was no problem with that.

Discussion number 2, the advisory committee is asked to comment on a proposal that if we were to use IBE for market access -- and this is an important part of this discussion topic -- when there is compelling reason not to, during the interim period, which I've defined as the next year, we're proposing that some conditions would apply.

The first is a new condition. The current guidance has 20 percent allowable difference in the test-to-reference ratio, the GMR as it was referred to. We're proposing that we change that to 15 percent.

We admit there's not a lot of data since last year, or a lot of scientific evidence to recommend that change. However, because of some of the behavior we've seen with this criterion in the data sets, we feel that if we're going to allow something into the marketplace with the IBE criterion, we'd like to have a better constraint on the mean-variance tradeoff that it currently allows.

We're also suggesting as another constraint the

subject-by-formulation interaction should be nonexistent if we're going to approve a product under IBE. If it's less than 0.15, we would conclude no significant interaction. Our dilemma is when that appears to be greater than 0.15. We have a question in our mind, is it real, is it due to the test product, or is it occurring by chance alone, and we have no way of determining that and we have some reservations about approving a product with a significant subject-by-formulation interaction.

Furthermore, we'd like to suggest that sponsors follow the recommendation that subjects should be heterogenous, taking into account age, sex, race factors, as appropriate, in conducting the studies in which they would like to gain market access using the IBE criterion. We feel that's necessary or it defeats the purpose of IBE, that is, in asking the question about variances and about subject-by-formulation interactions.

Discussion topic number 3 is a somewhat status quo question. You can see what it is on the slide. It's basically getting to the continuation of our recommendation to conduct replicate design studies for modified-release products and for highly variable drugs. We have no reason to suggest a change in this recommendation. About half of the products that we sent to you as new data were modified-release products. However, the subject population in

almost all those studies was a homogeneous population and not a heterogeneous one.

We feel that it's important to continue with this approach in the absence of data not going forward with it. With regard to replicate design studies, it provides us empirical evidence of any problems with ABE, if they exist. It continues to allow us to explore, as we have done, a systematic analysis of the subject-by-formulation interaction to resolve whether its frequency is enough to be of concern. And it allows us, on a case-by-case basis, to assess the clinical significance of differences in variance.

As I said, in order to do all this, we need a heterogenous population to maximize the information that we'll get from these studies in order to make any conclusions or extrapolations from these studies. It's difficult to do it using all males who happen to be young.

We'd like to get to a final destination with this individual bioequivalence and make a final decision to use it, not use it, when to use it, to allow market access. It's a significant scientific and public health issue. We want to be sure that we have the rationale to make the right decision.

So, we feel we need a larger database, recognizing that even one year won't provide us the entire

database we need to make the decision, but we need more actual examples, and we hope using a heterogeneous population, coupled with some simulation and other exercises that would allow us to come to a final resolution of this issue of its use.

Finally, our last discussion topic. We provided you with a research plan. We ask for a comment on it. The research plan is fairly comprehensive. We're not sure we have the manpower to accomplish it all. It's important, we feel, to have some priorities in this research plan. It hasn't changed substantially since 1999, but any comments the committee would have on priorities within that research plan would be beneficial to us.

So, that brings us to the agenda, and my introduction as to why we're here. What we'll hear at this point from the FDA speakers will be a presentation of the replicate designs in the ANDAs. This will be primarily new data that hasn't been presented before. We'll follow it up with a presentation of replicate design studies from the ANDA database, and finally we'll hear a presentation on the research plan.

Thanks very much.

DR. LEE: Thank you, Larry.

Does the committee understand what the marching order is?

We have two new guests joining us. Would you please identify yourselves.

DR. BOLTON: I'm Sanford Bolton.

DR. ZARIFFA: And I'm Nevine Zariffa from GlaxoSmithKline Pharmaceuticals.

DR. LEE: Thank you.

Mei-Ling, are you going to make the two presentations separately?

DR. CHEN: I will do that together. Well, good afternoon, everyone. As indicated, there are two parts in my talk, and for the benefit of new members on the advisory committee, I will briefly provide an overview of the background and the concepts of individual bioequivalence, and in the second part of my talk, I will then discuss the results of our statistical analysis for replicate design by bioequivalence studies, with the focus on NDAs in the FDA database.

As most of you know, the current regulatory approach for evaluation of bioequivalence has been based on the comparison of population means between products, and this is the so-called average bioequivalence approach. The agency has been interested in the individual bioequivalence because the new approach appears to offer several advantages over the use of average bioequivalence. The individual bioequivalence compares not only to population

means but also the variances between products. This approach considers subject-by-formulation interaction, which is believed to be an important factor in the assessment of switchability between products.

With an appropriate criterion, the individual bioequivalence can establish goalposts based on the reference variability, and this is particularly useful for highly variable drug products. The new approach also creates incentive for both innovators and generic sponsors to manufacture less variable products. Because of the emphasis on the assessment of subject-by-formulation interaction, this approach also encourages the use of a heterogeneous population in bioequivalence studies.

An important principle for individual bioequivalence assessment is based on the distance concept. The principle is to compare the distance between the test and the reference product with the distance between the test-reference and the reference formulations. So, for individual bioequivalence the test and the reference formulation have to be administered to the same individual. If we call this comparison an individual difference ratio, then the goal of bioequivalence demonstration would be to show that IDR is not substantially greater than 1.

So, based on the concept of distance ratio, the agency has developed the individual bioequivalence

criterion with a general form like this. It combines the average bioequivalence criterion with the variance terms, which is then normalized by the variance of the reference product.

So, the variance terms are subject-by-formulation interaction, sigma D squared, and difference within-subject variance between the test and the reference product. Those are sigma WT squared, and sigma WR squared. And theta I, on the right-hand side of the equation, is a bioequivalence limit specified by the regulatory agency.

Now, what is subject-by-formulation interaction? In simple language, it is a measure that tells us how similar or dissimilar each individual response to the test and the reference product. On this slide sigma D squared is the subject-by-formulation interaction variance component, and it's the variance of the individual mean differences between the test and the reference products. So, sigma BT and sigma BR are the between-subject standard deviation for the test and the reference product, respectively. Rho is the correlation coefficient between the individual means between the test and the reference products.

So, as you can see from this equation, there are two sources for subject-by-formulation interaction. It may come from the changes in between-subject variability

for the test and the reference formulation, and it may be due to the lack of correlation or congruence in individual means between the test and the formulation. Sigma D is zero only if sigma BR equals sigma BT, and rho equals one. So, based on this equation I would like to point out that sigma D is independent of the within-subject variability of the drug products.

Our experience so far has indicated that subject-by-formulation interaction does exist. In some cases we could identify the factors that contribute to the interaction, but in other cases we couldn't identify the factors or subgroups that caused the interaction.

This is an example that illustrates a subject-by-formulation interaction due to an age difference in the population. Two generic products versus a brand name drug, and as indicated on this slide, the test reference ratios for generic 1 are consistently higher in the elderly than for young people, and the phenomenon doesn't occur to generic 2. It's an age-based subject-by-formulation interaction, and the authors of this paper suspected that the higher serum levels of generic 1 might be due to the faster dissolution rate or absorption rate, which in turn saturated the hepatic enzymes in the elderly.

The second example came from the studies on a calcium channel blocking agent. The mean T/R ratios of

Cmax and Tmax in male subjects are significantly different from those in female subjects. This is a gender-based subject-by-formulation interaction, and the mechanism of this interaction has been postulated, which is related to a different release rate of the two formulations, and possible gender differences in metabolism and transport along the GI tract.

In the agency, we have seen other examples of gender-based subject-by-formulation interaction, but due to time constraints, I wouldn't be able to present them here.

How do we interpret the subject-by-formulation interaction? There are two approaches. One approach is to estimate the percentage of individuals whose average T/R ratios are outside a range of 80 to 125 percent. Another approach applies to the cases where the subject-by-formulation interaction arises due to the presence of subgroups that have different test-to-reference ratios from the rest of the population. I will explain this further in the next two slides.

This is a graphical representation of approach 1. The x axis is the sigma D value, and y axis represents percent of individuals with mean T/R ratios outside 80 to 125 percent. So, for example, if sigma D is .15, you see approximately 15 percent of the population individuals having their T/R ratios outside 80 to 125 percent. If

sigma D is .3, then approximately 46 percent of subjects would have their T/R ratios outside 80 to 125 percent. In this context if we consider 15 percent is a large proportion, then a sigma D value of .15 may be considered as a cutoff for a large subject-by-formulation interaction.

Bear in mind that this figure is constructed with the assumption that test-to-reference mean ratio is 1. So, if the T/R ratio deviates from 1, then the same sigma D value may imply more proportions of individuals having their T/R ratios outside 80 to 125 percent.

The second approach relates to the interaction where the formulations differ in a subgroup but not in the remaining subjects of the population. The x axis is the proportion of subjects in the subgroup, and y axis reflects a sigma D value. Each curve represents a fixed mean T/R ratio for the subgroup. So, ranging from 1.2 to 2. The larger the main T/R ratio, the higher the curve. As such, you can see sigma D value is a function of two factors: one, the proportion of subjects in a subgroup; and the second, the mean T/R ratios in that subgroup.

So, for example if I have 5 percent of the population having the T/R ratio of 2, you see the corresponding sigma D is .15. Similarly, if I have 25 percent of the population having a T/R ratio of 1.4, a sigma D is also .15. But interestingly, if you look at the

horizontal line for sigma D .15, this line across the board, then you see this line only intersects with those curves having T/R ratios greater than or equal to 1.4. In other words, if I have 50 percent of the population with a T/R ratio of 1.3, then the sigma D plateaus at .13 and it never reaches .15.

So, in this regard, using .15 as the cutoff for sigma D, it's not really strict. We have subgroups in the population, and it becomes important to choose the appropriate definition when we interpret sigma D.

Derived from the distance ratio, the individual bioequivalence equation ends up to have sigma WR in the denominator. This is interesting in that it actually represents a scaling approach where the bioequivalence criterion can be adjusted based on the variability of the reference product, and the reference scaling takes us away from the one-size-fits-all approach and offers flexible criteria for different classes of drugs.

One of the advantages of reference scaling is to widen the bioequivalence limit for highly variable drug products. It reduces the regulatory burden. In addition, the fact that sigma WR in the denominator is directly derived from the distance concept makes it sensible to have reference scaling using this criterion, rather than the average bioequivalence criterion.

The down side of this reference scaling approach is that we may unnecessarily tighten the bioequivalence limit for the drugs with low variability beyond a reasonable public health need. So, to correct this problem, the current guidance has recommended a mixed scaling approach. In other words, we set a regulatory limit for the within-subject variability, and that is called sigma W0. When the reference variability, sigma WR, is greater than sigma W0, we scale to the reference variance. When sigma WR is less than or equal to sigma W0, we scale to the constant variance.

As you can see from this equation, if the test variance is smaller than the reference variance, then it will be easier for the test product to pass the criterion. This provides an incentive for drug sponsors to manufacture less variable formulations.

In the meantime, it is possible to have a tradeoff between the mean and the variance, since both are in one equation. There was a concern in the past that the tradeoff in the possible -- also, reference scaling may allow a test product with a large average difference to enter the marketplace. To avoid this situation, the current guidance has recommended further constraint on the point estimate of geometric test-reference means, to be within 80 to 125 percent.

Turning to my second part of the talk -

DR. MOYE: Is it inappropriate to ask a question about the first part at this time? Do you really want to wait till the end of the second one?

DR. LEE: Is it a clarification?

DR. MOYE: I think it is.

DR. LEE: Please go ahead.

DR. MOYE: Perhaps you're using the word "interaction" differently than I'm used to. When I think of a subject-by-formulation interaction, I'm thinking that there is a dependent variable upon which the formulation can have an impact and the subject can have an impact. To my way of thinking, a subject formulation interaction is a modified effect of the formulation by subject. That is to say, the effect of the formulation differs from subject to subject. Is that what you mean?

DR. CHEN: Correct.

DR. MOYE: So, when you talk about a gender-modified subject-by-formulation, you're saying that the way the subject modifies the formulation's effect depends on gender.

DR. CHEN: No, that's not what I meant. It's actually, like you say, an interaction between the characteristics of the formulation and the individuals recruited in the study. So, the interactions actually

should be due to both factors: subjects and the formulations. But here what I illustrated is only on the subject side. In a way I could identify the factors based on the subjects. But I haven't really talked too much about the factors from formulations.

DR. MOYE: Well, I don't want to take too much time, but I did have that question.

DR. BENET: Vince, I'd like to ask a question?

DR. LEE: Sure, Les.

DR. BENET: Mei-Ling, I know that the reference product with the gender-based is on the market. But are there generic products also on the market of that reference?

DR. CHEN: Which one? The gender-based?

DR. BENET: The gender-based product.

DR. CHEN: The gender-based product. My understanding from the Generic Office was that that was a study presented by --

DR. BENET: No, no. That's not the question. The question is, are there generics on the market for that product which you have shown that the reference has a gender effect?

DR. CHEN: I think I don't know at this point.

DR. LESKO: I think I can answer that question, Les. The product Mei-Ling is talking about was never

approved for the market. However, there are generic diltiazem products approved in the marketplace. Calcium channel blockers. Sorry.

DR. BENET: What Larry has said is that the reference product is the innovator diltiazem product, and that there are generics on the market of diltiazem. That's what my question was. Is that the correct answer? Is that what Larry said?

DR. LESKO: Yes, it is. That's what I said.

DR. BENET: Thank you.

DR. LEE: Any other questions, since the floor is open?

(No response.)

DR. LEE: I assume not. Mei-Ling, please go on.

DR. CHEN: Now, turning my second part of the talk, I will show you some of the real data from replicate design bioequivalence studies.

For drug submissions, FDA previously collected 27 data sets. In addition, there were 28 data sets analyzed by the industry. After the publication of our final guidance, we have received 9 more studies from new drug applications and 13 more from ANDAs. So, in total there are 77 data sets with the replicate design studies.

Unfortunately, most of these data sets were

conducted in healthy, young male subjects, with a few exceptions of having females in the studies. Moreover, most studies in the FDA database have been performed on immediate-release dosage forms.

So, this slide gives you a snapshot of the old database. For the 27 FDA data sets, the frequencies of having a subject-by-formulation interaction of greater than .15 are approximately 20 percent for AUC, and 33 percent for Cmax. Because of the small sample size, some of these interactions did not show statistical significance. However, the confidence intervals with these interactions are wide, and so we couldn't really rule out the possibility of important subject-by-formulation interactions.

If we compare the with-subject standard deviations between the test and the reference product, using T/R ratio 1.2 as a cutoff, then the frequency for the test product having a higher within-subject variability is 33 percent for AUC, and 30 percent for Cmax.

It appears that similar results were obtained by the industry. However, their frequencies for subject-by-formulation interaction greater than .15 is a bit higher for Cmax. It's around 40 percent. These data have been previously discussed and presented at several meetings, so we will not discuss it here.

Our focus this afternoon will be on the new data set collected this year, and as shown on this slide, of the 9 studies from NDAs there are three modified-release, six immediate-release, and six highly variable drugs. Of the 13 studies from ANDAs, we have five modified-release, five immediate-release, and three slow-release, and three highly variable drugs. All the studies were conducted in healthy volunteers, and the sample size ranged from 17 to 93 subjects.

This slide summarizes the results of data analysis for three modified-release products. Bear in mind that all the analyses were conducted on the log transformed data, so the within-subject standard deviation on the log scale approximates the within-subject CV on the original scale.

So, for the three modified-release products, average bioequivalence and individual bioequivalence are in agreement with respect to the conclusion of bioequivalence. That means when the study passed ABE, it also passed IBE. When the study failed ABE, it also failed IBE. This is because there is no substantial difference in the within-subject variability between the test and the reference formulations, and there is no subject-by-formulation interaction in most cases, with the exception of data set number 3.

Data set number 3 is a study of an enteric coated dosage form, and the Cmax of this study failed ABE, average bioequivalence, because of the big difference in the T/R means. It also failed, I believe, because of the combination of the large mean difference in the subject-by-formulation interaction. A further analysis of the individual data has revealed that three subjects have their mean T/R ratios greater than 1.5, that I didn't present here.

This slide shows the immediate-release products. We actually have six IR products, and bioequivalence outcomes are also similar, using either IBE or ABE, with the exception of two AUCs in data set number two and AUC-infinity in data set number three.

As shown on this slide, data set number three has a sigma D .3, and it's a highly variable drug product, with the reduction in the within-subject variability, reference 40 percent and the test 35 percent. And also reference scaling, this study passed individual bioequivalence.

I have to talk about data set number 2. Data set number 2 has a big sigma D subject-by-formulation interaction for both AUC parameters, and therefore, this study passed average bioequivalence but failed individual bioequivalence. After further examination of individual

data, we found a subject with extremely low AUC and Cmax values on both replications of reference product. Some people may have a concern that the individual bioequivalence criterion is too sensitive for outliers. However, because of the use of replicate designs, we can actually check if the abnormal values come from the outliers. So, in this case the retest character of the replicate designs tells us that it's unlikely that this is due to outliers, because both values on the reference product are on the lower side. The question, then, is whether this subject represents a subgroup in the population who responds to the test and the reference differently.

I would like to switch gears to talk about the FDA contract studies. There are three studies --

DR. LEE: Mei-Ling, would you give us a quick summary?

DR. CHEN: Am I out of time?

DR. LEE: Yes, you're almost out of time. Because of the questions because of the questions Les asked I think.

(Laughter.)

DR. CHEN: Okay. I guess I have to summarize our contract studies. Ranitidine, metoprolol, and methylphenidate. I will discuss ranitidine and metoprolol

together because these two studies were performed in parallel to investigative effect of excipients on the bioavailability of drugs. Both studies compare the bioavailability of candidate drugs in sorbitol versus sucrose solution.

From the literature we know that ranitidine has low permeability, while metoprolol has high permeability. Regarding the two excipients, we know sorbitol has low solubility and permeability. It can increase the osmotic pressure in the gut and reduce the GI transit time. On the other hand, sucrose has high permeability.

The hypothesis was that the bioavailability of a low permeability drug such as ranitidine is more likely to be affected by an excipient such as sorbitol that reduces the GI transit time. And the subject-by-formulation interaction may occur when two syrup formulations contain different sweetening agents.

This is the result with ranitidine studies. You see sorbitol solutions produced lower bioaviability than the sucrose solution. While in the metoprolol study, the excipient effect has much less influence on the metoprolol levels.

Interestingly, we also found a subject-by-formulation interaction in the sorbitol ranitidine studies. In a way a reduction of between-subject variability from

sucrose to sorbitol resulted in a subject-by-formulation interaction, and sigma D is about .15. So, the point is that an excipient could also produce a subject-by-formulation interaction.

The last contract study is on methylphenidate. The study was conducted in the 1990s, and the test product was suspected to have poor quality and behave erratically in the clinics. It's a replicate design study, so we analyzed the data recently, using the individual bioequivalence approach.

The table shows the test product not only has a higher T/R ratio for Cmax and also has a higher within-subject variability. It also has a marginal subject-by-formulation interaction. With average bioequivalence, we passed the study, but with individual bioequivalence, we may have rejected the study.

Thank you very much.

DR. LEE: Thank you very much. Sorry to cut you off.

Quick question, Marv?

DR. MEYER: In your old database, page 9, you have 33 percent with a Cmax SxF greater than .15, and some other numbers. Does that imply you would reject or not pass 33 percent of the studies in the old data using IBE?

And then the second question is, under data set

2, page 11, where you have subject 9, to me that just looks like variability because you have, let's say for AUC you have a 727 and a 3680 for the reference given twice, and for the test there's close agreement. So, you have one high, one low. To me that doesn't look like a replicated subject-by-formulation interaction. That just looks like variability in the reference in that subject.

DR. CHEN: Let me answer the first question first. You're saying that if the subject-by-formulation interaction sigma D value is greater than .15, will we reject the study? Is that the question?

DR. MEYER: Yes.

DR. CHEN: No, not really. Because the current criterion is a composite equation, and sigma WD, subject-by-formulation interaction, is just one of the terms in that equation. So, we don't have a separate requirement to say sigma D needs to be less than .15 in order to pass the criterion. The current proposal in the guidance is to treat the whole criterion as a --

DR. MEYER: As a companion you also have your T/R WS SD, 30 percent of the database also had a value greater than 120. So, it seems like two of the components in your IBE are bad, so to speak.

Would you fail a number of those studies? Maybe not all 30 percent, but some percentage? Should they

have failed using IBE? Would they fail?

DR. CHEN: This is just to analyze all the data that we have at that point and to give us some appreciation of the performance for the test and the reference products in all the bioequivalence studies. We didn't use the IBE criterion for acceptance or rejection of those studies. Did I answer your question?

DR. LESKO: I'd like to try to answer that question because we have to be careful about an estimated value of sigma D being over .15, and as Mei-Ling showed, the estimated value of sigma D was over .15 in about 30-some percent of the studies. That does not necessarily mean that 30 percent of the studies had a subject-by-formulation interaction. Many of these studies are underpowered to accurately detect sigma D, and there's a possibility that many of those could be due to chance alone because of the low subject numbers in the studies.

So, one of our dilemmas is when we see these high values, and when we start to look at all of these cases, sometimes we can't find any mechanistic reason for subject-by-formulation interaction to have occurred. So, we have no way of sorting out, when the value is large, whether it's real, or whether it's occurred by chance alone because of an underpowered study.

DR. LEE: Kathleen?

DR. LAMBORN: I just want a clarification because if I interpreted what you're proposing to do, which is I think what the question is, the addition of the requirement that the subject-by-formulation interaction be less than .15 -- that's the estimated subject-by-formulation interaction. Right?

DR. LESKO: That's right.

DR. LAMBORN: And so under that criteria, these would have failed. Is that correct?

DR. LESKO: No, they wouldn't have failed because what we're proposing is if one wants to use the IBE criterion, it would have to meet this standard.

DR. LAMBORN: So, in other words, if they had used the IBE criterion for these studies, then these would have failed given the criteria of requiring less than or equal to 15 percent.

DR. LESKO: If the company had come in and said, I want to use a priori IBE for market access, then under those conditions, yes, that would be the case. But certainly there's another route to approval of those products.

DR. LAMBORN: No. I realize that. But we're just trying to understand how this data would have fallen if they had used IBE.

DR. LESKO: That's right.

DR. BENET: Vince, I have a question.

DR. LEE: Les, very briefly, please.

DR. BENET: Mei-Ling, in your analysis of the data set that you call data set 2 but in our tables are data set 3, where you showed that one subject had very high levels of AUC and Cmax, the implication is that that was the reason that this failed IBE.

Have you tested, if you delete that subject, whether the study would have passed IBE? My guess is it will not. Independent of that subject. Have you tested it?

DR. CHEN? I think I tested it, and if we were to delete the subject, this study would have passed IBE.

DR. BENET: Thank you.

DR. LEE: One final question.

DR. ENDRENYI: As suggested two years ago by the expert committee and also before this advisory committee, could various data sets be published on the Internet in detail?

DR. LEE: Who can answer that question?

DR. LESKO: The lawyers, I guess.

(Laughter.)

DR. LESKO: I'll have to check. I don't know. If it's an approved product, maybe. If it's not approved, maybe not.

DR. ENDRENYI: An earlier data set was published.

DR. LESKO: Yes, and that was an old data set, whereas this is a new data set for products that may be under review at the current time.

Vince, if I can make one more clarification for the committee, I think it's important to realize that when a subject-by-formulation interaction appears large, it isn't necessarily the test product that's producing it. It could be the reference product. We have to be careful to not assume that every time you see a large subject-by-formulation interaction, the test product is bad. In fact, in one of the data sets that Mei-Ling showed, which was number 3 on the NDA chart, that one with that large subject-by-formulation interaction, that was the reference product.

DR. LEE: Very well. I think that we should move on to hear about the ANDA situation from Dr. Patnaik. Les, are you available until 3:00?

DR. BENET: I'm available until 4:00.

DR. LEE: Great. Thank you.

DR. PATNAIK: Good afternoon. I am going to present some data from the ANDA side, and as you know, in your handouts there were 11 data sets. We have added two more to those data sets because we received those two

additional data sets, so we included that. So, I'll be presenting not 11, but I'll be presenting results from 13 data sets.

As Dr. Chen has already explained, I just put it down in simple words. With average bioequivalence, you evaluate the difference between the test and reference means, and you think that they should be within certain regulatory limits. So, here we are only looking at the difference between the two means.

As a contrast to ABE, the IBE looks at the differences in the mean and looks at the magnitude of the subject-by-formulation interaction and the differences in the within-subject variances. Then you normalize it with the reference variance, within-subject variance, or the regulatory within-subject variance, whichever applies to the reference variability.

So, if it is more than .2, you reference scale. If it's less than .2, you have the regulatory constant within-subject variance to normalize with. This left-hand side must be less than or equal to a regulatory bioequivalence limit. So, it has got three components, as opposed to the single component in the average bioequivalence.

Now, I'll just give you a summary of the studies. These studies were submitted for the approval of

generic drugs. There are 13 studies which we will be discussing and the study designs more or less are two-treatment, two-sequence, and four-period crossover designs. These are the two designs, which have been used for these 13 studies.

The number of subjects, starting from the low number of 16 to about 60 subjects, and they are usually a controlled population, mostly young, healthy male subjects.

And there are several types of dosage form which have been studied, immediate-release and modified-release. They are all solid oral dosage forms, and one suspension and one suppository. What I will talk about is only the parent drug. We will not include metabolites.

These are the three bioequivalence measures: the two AUCs and the Cmax.

This is like a global result. What I've done because we're talking about average BE and individual BE, so I just gave a very global view of how much of these data sets pass or fail individual and average bioequivalence. In this column average bioequivalence will pass, and in this column average bioequivalence will fail. In this column individual bioequivalence will pass and individual bioequivalences fail.

As you can see, 11 out of 13 data sets and 12

out of 13 data sets passed both average and individual. Only 2 out of 13 and 1 out of 13 for the AUCs fail IBE while passing average BE.

On the other hand, one data set out of 13 only failed Cmax. None of them failed AUC. None failed both of them, either average or individual. So, all of them passed. None of them failed both criteria.

The lower part shows the range of numbers and data which has been received which has been analyzed, and the results show that. The mean ratio is from 11 percent lower for the test, to over 4 percent higher for the test compared to reference. And the within-subject standard deviation varies from 6 percent CV to over 40 percent CV for AUC, and similar value for AUC-infinity. But for Cmax it had never gone below 11 percent and from 11 percent to about 45 percent is within-subject variability for Cmax.

In terms of ratio of the variances, it varies from 12 percent -- the test variability is 12 percent lower to about 56 percent higher than the reference for test, and the same thing, very large, about 23 percent lower than the reference to about 55 percent higher than the reference.

Here as you can see, the ratios are all over the place, from about 70 percent lower than the reference to about 35 percent higher than the reference. So, it's a very broad range of ratios.

And the subject-by-formulation range is having no subject-by-formulation to a maximum of .2, and that to a very few data sets. So, this is the global picture of the whole 13 data sets.

The next slide. I have just put everything on a bar graph so it's very easy to understand, and maybe it will give a clearer picture. This is the within-subject standard deviation of the reference product. The upper panel is for AUC, the lower panel is for Cmax.

The y axis is the variance term, within-subject standard deviation, and these are the data set numbers on the x axis, and these are what kind of products. I classified them into higher product, one suspension, two slow-releasing products, and five of them are extended-release product, and one is a suppository.

Here you can see that, as we said earlier, anything 20 percent or higher in within-subject variability of the reference, the criterion asked for reference scaling. When the standard deviation is lower than 20 percent, we have to do the constant scaling. So, in this case we have one, two, three, four, five, six, seven, eight. Eight data sets will be reference scaled, and five data sets will be constant scaled.

As you can see, we're talking about highly variable drugs. We call them highly variable when the

within-subject standard deviation on CV is more than 30 percent, as Dr. Midha said. So, we have only one, two, and three products, two IR's, and one suppository can be considered a highly variable drug product. This is for AUC.

But for Cmax, under the same reference, we have gotten the same eight data sets will require reference scaling, and another five data sets will be constant scaled. In this case, we have only two data sets which can be considered as highly variable for Cmax.

Now, this looks complicated, but it's pretty simple. What I've done is here in the three panels, I have put the test-reference geometric mean ratio in the top panel, the test reference within-subject standard deviation ratio in the middle panel, and the subject-by-formulation in the lowest panel.

Now, these are the drug numbers, data set number. The left-hand side y axis is the log transformed test-reference ratio, and the right-hand side is the linear, showing 1.04, so the ratio is 1.04.

The reason I did that, when the ratio is below 1, you see a negative number, so anything the bar shows below 1.0 or below 0 in the log scale, the test is lower than the reference, and in the upper part the test is higher than the reference.

So, one can see here that data set number 2 or drug number 2 and drug number 6 have got around .88 ratio. Test and reference is .88, about 12 or 13 percent lower AUC than the reference. So, also the drug number 2.

Correspondingly, one can see for the drug number 2 the test-reference ratio for the variances is about 11 to 12 percent. This test is 12 percent higher than the reference in terms of variability.

There are only two. Number 6 has got a very high ratio. It's about 50 percent of the reference.

Plus these two slow-releasing products, number 5 and number 10, the arrows are showing the ratio of about more than 55 percent of the reference. So, there are large differences in the within-subject variability for these two drug products, and also this number 6.

Now, the yellow and red shows the failing of IBE criterion. So, there are two drugs, number 4 and number 6, failed IBE criteria. Now, they failed here, as you can see, because of the large differences in the geometric mean ratio, the large difference in the within-subject variance, but there is absence of subject-by-formulation. So, these two contributed to the failure of IBE.

In this case, where number 4 is failing, it is because although it is only 4 percent difference in the

mean, it has got about 22 percent or 23 percent higher difference in the variance, but it has got a large sigma D, or within-subject variance. So, these two contributed to the failure of number 4.

So, this is just a comprehensive picture of what is happening between the three components for the same drug product.

This one is for Cmax and you can see there are no red bars here, so everything passes IBE for Cmax, and here also there is quite a difference. About 16 percent higher you see in data set number 7 in the test-reference ratio for the mean. And you have one drug product, drug number 2, which shows large subject-by-formulation interaction, but it doesn't flunk IBE because the ratio is not that much. Test-reference variance ratio is not that much. And the test-reference mean ratio is also not very large. So, none of them fail IBE.

To come into the specific examples very quickly, we talked about that one drug which fails average BE. Drug number 2 is an IR product which failed for Cmax. N is equal to 55. And the difference is about 12 percent in the means, and it falls just marginally, and if they would have taken more subjects, probably it would have crossed the 80 percent mark. It passes IBE in spite of this 13 percent difference, as well as it has got a .2 as a

subject-by-formulation interaction. It is a highly variable drug, but the ratio of the variability is very comparable, not very large. So, here the reference scaling really helped to pass this IBE. It would also have passed ABE with a couple of more subjects. So, this is why it fails ABE but passes IBE, mostly for reference scaling.

The second example is to pass IBE. Drug number 1 is an IR product with 29 sample size, but it fails IBE. Drug number 4 is the immediate-release. N is equal to about 59. It's for AUC 0 to T.

In one case, these are the two things just comparable. I just put it down to see a comparable observation. They're almost same point estimates, 2 to 3 percent in means. Here in one case there is no subject-by-formulation interaction for drug number 1, but for drug number 4, .2 is the subject-by-formulation interaction. Like drug number 1, it has got very low and similar within-subject variance.

In one case you have got a 36 percent difference in the standard deviation difference, within-subject standard deviation mean, as well as the difference, and here it is 23 percent. So, what is happening, that this 23 percent difference in the within-subject variance, higher, and the presence of this subject-by-formulation interaction, even though it's .4 percent difference in the

mean, allows it to fail IBE criteria.

So, these are the behavior performance of the almost similar type of data, showing one passing and the other one failing, particularly for this large subject-by-formulation interaction.

Number 3, which is important, is that it passes IBE. It's a suppository with 57 sample size. Drug number 3 is the last bar in the graph. It's failing IBE. Drug number 6, with extended-release product with 27 sample size. Here point estimate is just like on the dot, which passes. There is no subject-by-formulation in either case. This is a highly variable drug, but here you see the reference formulation has got much lower variability than the test formulation, and that is why the ratio is about 50 percent higher. That makes it to fail.

So, there are two different performances as compared to 2 and 3.

When I looked at these two high subject-by-formulation interaction in this case, that is drug number 2 and drug number 4, which has got a subject-by-formulation interaction, I just wanted to look at each of those data sets. I'll go very, very quickly.

There are three subjects which stand out as abnormal data. Subject number 13, subject number 53, and subject number 38. In one case for the reference, this is

the sequence of administration in two different periods. The reference is very consistently low and the test is very consistently high. And in this case also it's also dissimilar.

It doesn't fall in a big pattern in the sense that in one case the test is higher than reference. In this case the reference is higher than the test, and in this case the test is higher than the reference.

Now, once we look at those things, and if you want to look at which one is responsible for this, to a certain extent these affect very marginally, but this subject affects that subject-by-formulation very dramatically.

The other example is for the Cmax. As I said to you, same thing. You have the two tests showing higher than the reference. But in this case there's one treatment out of test and one treatment out of reference are showing abnormal value. Now, which one is the outlier? Is this value an outlier with respect to this, or this is an outlier with respect to this? It's very difficult to say. But this is pretty consistent.

And here also I found out there's marginal effect on removing the sigma D, but there is very dramatic effects of removing sigma D for this.

So, my concluding remark is this, that we have

to think about the IBE criterion as an aggregate criterion. We cannot separate the components out. Just to evaluate the performance of the criterion at least. So, the combination of those three parameters, they determine the outcome.

Scaling approaches were seen, and I've shown to you, are particularly helpful for highly variable drugs with very large within-subject variance.

Analysis of the data showed that important subject-by-formulation interaction occurred due to very, very few subjects. At least, it's a very limited number of data sets. But the reliability and the possible cause of such observed interactions need to be carefully investigated. Why it occurs, I don't know. It's very difficult for me to say.

The studies we've received thus far, finally, during this period have utilized controlled populations. We have talked about this. The frequency of occurrence of important subject-by-formulation interaction and the utility of this approach I'm pretty sure will be better understood or evaluated as more BE studies using heterogeneous general populations become available to the agency.

Thank you very much.

DR. LEE: Thank you. I'm going to hold the

questions and go right to Dr. Benet, who has been asked to speak on behalf of the scientific community. Les, are you there?

DR. BENET: I'm here.

DR. LEE: Please go ahead.

DR. BENET: My slides will always come delayed, so I'll just assume they're up there.

Thank you for the invitation to talk, and I'm sorry that I can't be there in person today, but I want you to know that you all look very good on televison and I'm enjoying looking at you.

(Laughter.)

DR. BENET: I was asked to make this talk and to select a title prior to seeing the data and information that was provided in the book. So, I had to select a title not knowing what I was going to look at, so I selected this title, and I will talk about that briefly.

I believe that this is the opinion of the scientific community, but it's a group that would be generally favorable to IBE, and that group would say that individual bioequivalence is a promising, clinically relevant method that should theoretically provide further confidence to clinicians and patients that generic drug products are indeed equivalent in an individual patient. I think that's a lofty goal and it would be nice if it was

true.

On the next slide, I believe that this is the opinion of everybody. Even today, considering the studies summarized and analyzed by the FDA, the data is inadequate to validate the theoretical approach and provide confidence to the scientific community that the methodology required and the expense entailed are justified. I certainly think that we heard that during the open session.

The next slide I believe would be the opinion of the majority of the scientific community, and that would be that at this time individual bioequivalence still remains a theoretical solution to solve a theoretical clinical problem. We have no evidence that we have a clinical problem, either a safety or an efficacy issue, and we have no evidence that if we have the problem, that individual bioequivalence will solve the problem.

So, that meets the criteria of my title, selected prior to seeing the data.

On my next slide, I have a new title, and that's the title now that I've seen the data. That title is, "Opinions and Recommendations of the Former Chair of the FDA Expert Panel on Individual Bioequivalence."

My overall position is, we don't have a problem with bioequivalence at present, and there is no issue that has been raised that creates a problem that should be of

concern to the scientific community in terms of safety and efficacy.

I have maintained for many years that the present plus 25 percent/minus 20 percent average bioequivalence equivalence criteria are extremely tight and that in fact these criteria have sufficiently served us to make sure that we don't have bioequivalence problems for approved drugs.

Now, one way that we look at problems for approved drugs is to see phase IV reports, and I think there is a lack of problems based on this issue. But in reality there have been much more data that has been available that we have never seen because this is a huge financial issue and the innovators have spent tremendous amounts of money and time attempting to show that approved generic products are not equivalent and that they have potential for safety and efficacy issues.

So, these prospective studies, usually carried out in special population subsets, have been carried out to attempt to demonstrate lack of equivalence for approved generics, and of course to demonstrate efficacy and safety issues, but you never see any of these results because none of the studies come out the way that the sponsor would like them to be.

Now, I'm aware of these studies because I've

run a bunch of them, and others are aware of them. And what we know is that we have tested prospectively the present criteria numerous times, and there's no issue.

So, on the next slide, I think it's important for us to look at what we are trying to solve, and at least two of these issues have been covered, but the third has not.

The first is the issue that for wide therapeutic index, highly variable drugs, we should not have to study an excessive number of patients to prove that two equivalent products meet preset, one-size-fits-all statistical criteria. And this is part of the driving force of the agency in looking for new approaches that would allow us to approve drugs without studying them in an unreasonable number of subjects where that is required by the present criteria.

The second issue we are trying to solve came about as we were attempting to look at this but was also focused very strongly by the narrow therapeutic index issue drug raised by the brand name industry. For all drugs, but particularly for NTI drugs, a practitioner may transfer a patient from one drug product to another and be assured of comparable safety and efficacy, that is, switchability. So, this is another one of our goals.

And we have a third goal that has not been

discussed at all in this advisory committee and that is to give patients and clinicians confidence that a generic equivalent, approved by the regulatory authorities, will yield the same outcome as the innovator product. Not to prove that it does, but to give them confidence that it does. And this is one of our major problems.

Now, I get invited to many conferences that are clinically based, and I am the representative individual that says that the generic product works just as well. Oftentimes I go when the FDA has refused to go, and the FDA refuses to go because most of these clinical conferences are sponsored by the brand name industry for a large fraction of the funding, and the FDA makes the position that this is potentially a setup or a conflict of interest. So, Les Benet gets invited. So, I go to those meetings and I hear all of the clinicians and their very strong concerns about the present criteria and future criteria that we are discussing.

So, in my mind, one of the most important things that we have to do is not only scientifically and statistically prove that these products are equivalent, but we have to have assurance of the clinical community that then is translated to patients that in fact drugs will work the same when they are a generic.

Let's go to the next slide which is discussion

of the subject-by-formulation interaction term.

My position is switchability is not a problem for approved generics at the present time under the average bioequivalence criteria. This is based on the fact of the statement I made earlier that our present criteria are sufficiently strict in terms of approval, and in fact, we have no problem. We do have anecdotal reports, and maybe those anecdotal reports are related to a particular switchability, but prospectively those kinds of issues have never been able to be quantitated and demonstrated by the brand name industry. I don't think we have a problem. I think what we have done is create a problem for ourselves by suggesting that we have products on the market that aren't switchable.

Now, I'd like to take two examples from the data that Mei-Ling presented. One is the one I asked the question about. This is diltiazem. The gender effect is on the innovator product, and that gender effect is real and it's a 30 percent difference. There are generics on the market. They probably don't have that gender effect. That was the question that I asked Mei-Ling and Larry. And why? Because we know very well that it is extremely hard to show any difference related to a 30 percent change in plasma concentration that's going to translate into any relevant pharmacodynamic response, both safety and

efficacy, and especially for a drug like this that is not a narrow therapeutic index drug.

Now, I'll go on to the third point, and that is, I noted in my meeting yesterday in a telephone conference with the group at the FDA, that when we looked at the new data NDAs, that the high subject-by-formulation interaction terms occur when the reference within-standard deviation is greater than the within-standard deviation for the test. And I particularly asked Mei-Ling if she took out the one subject, would they pass, and my bet is they wouldn't. Mei-Ling says yes, but I would like to see that.

Now, I am very concerned that we have a criterion that basically will fail a generic product because an innovator has high variability, and that's what we have. We have a situation where a product can fail subject-by-formulation interaction because the innovator product has less variability than the reference product.

Now, theoretically we've solved that problem by putting into the equation the difference between the within-subject test minus the within-subject reference. So, there is supposed to be an advantage for the test product to have less variability than the reference. But it is my contention that in fact this turns out to be a negative in terms of sigma D.

I do not agree with Mei-Ling's suggestion that

the equation for sigma D has nothing to do with within-subject variability. The equation does not include any terms for within-subject variability. But if we have a reference product that is highly variable and a test product that is not highly variable, it is hard for me to see how you will not have a sigma D that is not influenced by this difference.

And I think this is one of the major problems of the present approach, that in fact you can have a sigma D that is very large because you've got a better generic product, and I think this was demonstrated by some of the other individuals. I was able to hear some of the presentations during the open forum, and I was able to see Professor Endrenyi's presentation. My view is that this is a problem that is not an advantage as it's supposed to be. It's a disadvantage.

DR. LEE: Excuse me, Les.

DR. BENET: Now, I believe in the fourth point it is not reasonable -- and it's not going to happen -- to expect that sponsors will use subjects representative of the general population in their IBE studies. And I don't think we can legislate it appropriately. So, I think we're always going to see some kind of excuses, and if I'm a sponsor, I'm going to do the best I can to have the population be as conforming to a standard as possible.

My view is that the subject-by-formulation interaction term is a red herring. There's nothing valuable about it and we ought to get rid of it because it doesn't solve any of the problems related to the statistics. It creates problems.

When I look down the list of issues where there was a sigma D failure by the IBE, in general those were in situations where you saw the test within-variance being less than the subject within-variance. Now, I'm talking about when there's a difference between passing ABE and passing IBE. You pass ABE, but you fail IBE.

My conclusion is I see nothing to suggest that we have anything useful by including the subject-by-formulation interaction term. I think there's no good data to suggest it's useful and I think we ought to get rid of it.

DR. LEE: Les?

DR. BENET: Next slide.

DR. LEE: Les, I think that we have to sum up.

DR. BENET: No, no. Mei-Ling went forever.

(Laughter.)

DR. BENET: IBE, ignoring subject-by-formulation, should allow sponsors to gain approval for highly variable, wide therapeutic index drugs without using an excessive number of test subjects. This is the reason

we should be doing IBE, for this purpose only. And as Mei-Ling and her colleagues have shown in the paper in 2000, it really only becomes useful if you've got a CV greater than 50 percent.

So, my preliminary recommendation is, on the next slide, that sponsors may seek bioequivalence approval using either ABE or IBE and with subject-by-formulation interaction deleted from the equation. If an IBE study is carried out and the test product fails, the data or a subset of the data may not be reanalyzed by ABE for approval.

Now, we have a perception problem that went to the third issue, as seen on the next slide. One of those perceptions is with IBE, that we could possibly allow approval of test products where mean bioavailability may fall outside of the 80 to 125 percent for the reference.

But we also have a perception problem with ABE because we now have a situation that if the products really have reasonable coefficients of variation and they differ and they really do differ, even between 10 and 20 percent, sponsors can get those products approved by just adding, adding, adding subjects. I don't think that is a useful approach.

Now, on the next slide, in March of 1998, I proposed formally this point estimate criteria. And I

believe that we need a point estimate criteria. It has nothing to do with statistics. It has to do with credibility of the process. I do not believe that we can go to clinicians and say, these two products on average differed by 30 percent, but they passed our criteria. Therefore, you should prescribe them and you should have confidence. I don't believe they're going to have that confidence, and that was the reason I suggested initially that we need a point estimate criteria.

I now in my final recommendation believe that we should have a point estimate criteria both for ABE and for IBE and that it should be plus or minus 15 percent, as the agency is proposing, for AUC, but higher for Cmax, and that consideration should be given for narrower point estimate criteria for NTI drugs because this is the perception problem.

In my view -- and we have the data that show it -- these are not problems. All the products pass these kinds of things. There are one or two exceptions that don't pass, and so I think it is important. I disagree with Laszlo. I disagree with Kamal. I do believe from a perception point of view that it is important to give the clinical and the patient community confidence that these products do not differ in the means, which is what they understand. They will not understand the statistics.

Thank you.

DR. LEE: Thank you very much, Les. I appreciate your insight, and I think that maybe the committee is ready to vote.

(Laughter.)

DR. LEE: Larry, you have a question.

DR. LESKO: I was going to make a couple of comments, if I can.

DR. LEE: Please.

DR. LESKO: I think Les put a lot of stuff on the table. I can't possibly sort through all of the things he suggested, some of which would involve some of the new methodology.

However, I just wanted to make a statement that the goal of an approval of a generic drug is to approve a product that's similar to a reference product. Similar means it's not going to be better or it's not going to be worse. A patient being switched from a reference product to a test product should expect to have the same safety and efficacy. So, that's just a general statement.

The other thing is, putting aside the subject-by-formulation interaction value of .15 for the moment, we do have data, and Mei-Ling presented some of this with our calcium channel blocker. But we have some other data. Is it overwhelming? No, but we have other examples where

there are some subgroup differences in the bioequivalence between the test and reference product when we look at it from a male subject and a female subject standpoint.

For example, one might have a test product that is 35 percent higher than a reference product, and when you begin to look at that, you see that much of that increase in bioavailability is due to the contribution from the male subjects as opposed to the female, or something like that. So, the differences in the bioavailability of the products sometimes will differ with identifiable characteristics of the subjects.

I guess the question that does raise, however, is, are those subgroup differences that we see maybe not necessarily important or unimportant, but are they best addressed through the individual bioequivalence paradigm, in other words, a subgroup difference? And the one Mei-Ling had was identified through a nonreplicated two-way crossover study. So, these things can be identified in alternative ways, but I think they do exist and I think we should pay attention to them and try to get more information on them.

DR. LEE: Thank you.

Let us go to the final presentation before the break, and we do have lots of things to think about. Dr. Machado is going to show us the plan of the FDA.

DR. MACHADO: Good afternoon, everybody. My task is to briefly describe the research plan for bioequivalence criteria, and you have a copy in your packet.

In terms of pertinent background, you know about the guidances that were issued in October of 2000 and in January of 2001. At the Advisory Committee for Pharmaceutical Science, in September of 1999 we discussed the FDA plans for further research and projects associated with the use of ABE and IBE criteria.

The advisory committee endorsed plans for furthering mechanistic understanding of using the IBE criteria, endorsed plans for conducting clinical pharmacology studies, and looking at the influence of outliers on the subject-by-formulation interaction.

At the same time, the committee requested creation of a research document to guide activities during an interim period, and a draft document was sent shortly afterwards for review by the expert panel that was led by Dr. Benet at the time. The draft research document was modified by our Population and Individual Bioequivalence Working Group, and this draft became ready in April of 2000. That is in fact the version that you have in the packet just with the date changed.

Now, the overview of the research program. The

overall goal is to provide information to support final regulatory decisions regarding criteria for comparing bioavailability and bioequivalence studies. The research plan has three components. First of all, to further investigate the criteria for bioequivalence comparisons. Second, to study issues related to data analysis and the statistical methodology. And third, to gain greater mechanistic understanding of any mean and variance differences that might be found between the test and reference products and also subject-by-formulation interactions.

Now, replicate design studies conducted by drug sponsors will be the major source of data for our evaluations. We're beginning to be ready to do some computer simulations to beef up our working data set, but right now the major source is from sponsors.

The general guidance recommends replicate designs for highly variable drugs and modified-release dosage forms. And that's been much discussed this afternoon.

Now, as far as criteria for bioequivalence comparisons, in our plan we plan to determine which criteria are appropriate for particular types of regulatory submissions, INDs, NDAs, ANDAs. And for the moment, we are using the ABE criterion for regulatory purposes. As you've

seen, we are analyzing replicate design data sets as we receive them -- 25 so far -- and interpreting these in light of the recommendations in the guidance. And the analyses we're doing will add to our knowledge base for evaluating the performance of the statistical and the other approaches and provide support for future decision making.

Also in the plan was that we would identify and evaluate clinically important test-to-reference differences in within- and between-subject variances and evaluate subject-by-formulation interactions. We will assess the importance and impact of the mean/variance tradeoff that's been commented on and look at other outcomes based on selected disaggregate criteria. We'll also study the discontinuity aspect of the individual bioequivalence method and possible resolution.

As far as the data analysis and the statistical methodology project, our main task is to evaluate the methods as laid out in the guidances, and we believe, after many years of work, that they're valid and reasonable. I should say we are open to new approaches, but we see our main task really to evaluate the characteristics of what we have and we've not seen anything so far that would make us abandon those methodologies.

Now, our objectives are to assess the estimation methods in the presence of missing data. That's

an important topic that hasn't been touched on. To further assess the statistical properties of estimates of parameters that we're most interested in, and to assess the impact of apparent outlier data on the properties of the aggregate individual bioequivalence criterion.

Other issues that we intend to work on are to monitor and assess possible carryover effects using data from replicate designs, but that depends on the actual drug being studied. And a fairly important objective is to assess the proper numbers of subjects and good study designs for heterogeneous populations that include both genders, possibly different ethnic groups and different age ranges, and consider what information we can draw from these studies.

Now, the third project is the mechanistic understanding. If we do find differences in means and within-subject variances, what might this arise from? And this would be done for the highly variable and modified-release drug products. Also, the subject-by-formulation interaction needs to be well studied in terms of mechanism.

So, our focus for the immediate future is to continue evaluation of the data from the replicate design studies as we're receiving them. In addition to the database that's accumulating, the interim period has about another year to go, and we've received 25 studies.

Possibly there will be another 25 coming in over the next year, and that isn't a huge database.

Now, what we seriously will consider is addressing the design issues, numbers of subjects, behavior characteristics of the various statistics. We can do this by computer simulation studies, and this will be based on the information in the databases to get realistic sets of parameters.

We will be evaluating the impact of possibly changing the constraint on the mean difference or imposing a constraint on the subject-by-formulation interaction to study the performance of the individual bioequivalence approach.

And last, but definitely not least, is we shall respond to the recommendations of the advisory committee.

So, finally, to summarize where we are, we see ourselves in the phase of evaluation using these data sets and simulations to understand the performance of the estimation methods for the remainder of the interim period. Just as a note, some of the issues that were laid out in the research plan, which was not changed since April of 2000, were in fact thought about, worked on, and addressed in the guidance.

Thank you.

DR. LEE: Thank you very much, Stella.

There is time in the discussion period to provide you with input on the plan you proposed. So, I'm going to suggest we hold the questions and go to a short break. Please, we will reconvene at about 3 o'clock.

In the meantime, what I'd like to do is to ask to have the four issues shown on the screen, and I think that the first issue is quite straightforward. We'd like to spend lots of time on the second issue, the third, and the fourth.

So, with that thought in mind, please take a break and come back at 3 o'clock.

(Recess.)

DR. LEE: I'd like to reconvene the meeting. I think this is where the fun begins.

Larry has posed four issues to the advisory committee, and those will be shown on the screen momentarily.

I would like to inform the group that we have the benefit of participation from several guests at the top of the table, and I invite them to contribute as they see fit.

Dr. Marv Meyer has kindly agreed to state some positions for the committee to react to, and I would just like to begin by introducing discussion topic number 1. Les, are you there?

although the data seem to be coming in slowly. So, I would be kind of neutral on another year, but I think that we should definitely continue to use average bioavailability for market access unless a company wants to come in and make a case for IBE. Highly variable, in my view, is the only reason to use IBE at the present time.

DR. LEE: So, Art is ready to have a counterpoint.

DR. KIBBE: I don't know whether I'm counterpointing, but I think we could take the first topic and put a period near the end of the second line. "Unless there is a compelling reason not to," period, and cross out the rest.

I'm not excited about the thought of converting all future submissions to IBE. I don't think that there's justification for that. I think there might be justification for allowing some submissions to follow an IBE methodology.

There are a couple of things that come to mind that we haven't talked about yet, and I want to just put on the table. If the agency goes forward and says that ABE is no longer acceptable and IBE is the method, is it in fact saying that the vast history that we've used ABE to get products on the market are not acceptable and how much retrofitting are we going to have to do? If you remember

all of the committee work to get pre-'36 drugs and all the OTCs reviewed, I don't know whether we really need to go back and do any of that. I think we are implying that we might need to if we go 100 percent for IBE.

DR. LEE: Jurgen, do you have an opinion?

DR. VENITZ: I guess I'm one of the scientific community that Dr. Benet quoted as thinking of this as a solution to a theoretical problem. So, I have no problems in saying that the current system, ABE as it is, works.

I'm very much like Marv. I'm neutral about collecting additional data. I'm not sure that additional data would help us to make a better decision next year or two years from now than it would be to do now.

DR. LEE: Thank you.

Dr. Barr has his hand up first.

DR. BARR: Yes. I'd like to take a different opinion I think. First of all, I take issue I think that we don't have a problem in terms of subject-treatment interaction. I think that we are just beginning to appreciate the extent of the problem and we don't know at this point in time how best to study that. Whether the aggregate approach is the best approach or whether an alternative approach is best, I don't know.

I'm also concerned about the aggregate approach, looking at too many things all at the same time

to the point you're not sure what the result is when you get it. So, we've attempted to go to the aggregate approach because it penalizes a company if they have to pass three studies, for example, or three criteria rather than one, and that was the reason that ultimately that I think the committee went to the aggregate approach. But we're finding that in collapsing all of that information into one number, that that may not be the appropriate way to go.

But on the other hand, to throw out, again, the baby with the bath water, like we did when we went to the 75/75 rule a long time ago, in which we had a method to look at individual bioequivalence, but it wasn't statistically sound, so we threw it out completely. And we now have no way of looking for subsets. And to go back and make that same mistake again for statistical reasons doesn't make sense to me.

DR. LEE: So, you said it's premature to throw it out.

DR. BARR: Oh, I think it's premature to throw it out and not look at ways to look at the subject-formulation interaction or the subset problem.

The problem of highly variable drugs we've already addressed in at least three meetings that I'm aware of in the past, and we always came to the conclusion that

we ought to treat highly variable drugs different than we do normal drugs that aren't as variable and extending the goalposts and allowing those to get through. So, that solution is already there. We don't have to have IBE in order to do that. We do need to address it. But I think that the real issue is how best to look for real subsets.

There are drugs that have been recently withdrawn, for example, a cyclosporine, in which people who eat had different bioequivalence for one product than they did with another. That would be a subset. If people want to look to phase IV kinds of withdrawals, they are out there.

People say that we don't know whether there are any subject-treatment interactions. I recently did a study that wasn't intended to find a subject-treatment interaction that found that there was a significant treatment interaction for levothyroxine products. I went back and found other studies that found the same thing, but ignored it by looking at an alternative way of evaluating it simply because they didn't want to see that. And I think that these things have not been seen because they haven't been looked for.

We certainly wouldn't see the gender effect because most of the studies in the past have been done only in males.

So, I think we ought to be sure not to make the same mistake of throwing that out again and not looking at it carefully.

DR. MOYE: But to pick up on that last point, it seems to me this is a highly unusual way to look for a demographic subgroup effect. There are established stat methodologies which allow you to specifically look for subgroup-treatment interactions, and they don't use this approach. It seems to me this approach is a new novel way to work out an effect that perhaps is not of the greatest interest after all. If we're really looking for a demographic, be it ethnic or be it gender or be it age, treatment interaction, then there are other ways to go with more established methodology with clearer track records than this.

So, I'm all for the development of stat methodology, but I suppose I'm just not clear on what problem, what question this particular stat methodology is trying to address. If it's trying to address an interaction which is a subgroup interaction, then I am in favor of rejecting this for the more traditional, standard approaches for looking at interactions.

DR. LEE: Larry, would you like to respond to that?

DR. LESKO: With regard to Bill's comments

about cyclosporine, I think we have a situation there where the problem with the formulation in a physical environment was the issue that was a problem there. That is to say, there was not necessarily an interaction between a subject's physiology and the absorption of the drug as much as there was a problem between the formulation of the product and when you admix it with a food environment, represented by juices basically. So, I'm not sure that's by definition a subject-by-formulation interaction as much as it's a food effect on bioavailability issue.

With Lemuel's comment, I think we sort of moved from an individual subject-by-formulation interaction idea where the methodology looks for a fraction of people in the test population that might demonstrate some unusual behavior with regard to either a test or a reference formulation. We sort of moved from that, which was the original concept of the IBE criterion, to the subgroup effect. And I think we did that because it's very easy to identify the subgroup in these studies where there's a retrospective analysis.

So, it isn't the intent of the approach to look for subgroup differences because I tend to agree with you, there are better ways to do that. In fact, the differences that Mei-Ling showed with the calcium channel blockers and with the verapamil came from non-replicated studies. And

one could do that under the current standard of average bioequivalence. But those are the known identifiers that might identify a population who would interact differently with the test and reference formulation.

What this approach was intended to do was to look for other factors that may be related to the range of physiological variables within a subject's gastrointestinal tract that somehow might distinguish between a test product and a reference product in a way we don't understand, although we can hypothesize on it, but we haven't really explored. That was sort of the difference between the subgroup and the individual.

DR. LEE: Kathleen?

DR. LAMBORN: I had sort of two thoughts. One is on the subject-by-formulation interaction and that criterion that was proposed of the 15 percent. My concern is, on the one hand, Les I think is quite right, if we allow things either in terms of the estimated ratio or the estimated subject-by-formulation interaction to be too large, even if they could be due to chance, we're going to have a perception problem which needs to be addressed.

On the other hand, putting in a criterion which says we're going to estimate this and it must be less than 15 percent and then you look at those that we think are really equivalent and you see that because we know

statistically that there's a great deal of variability in those estimates, given the sample sizes that we're talking about using, we're going to fail an awful lot of cases. And if we assume that in most cases they are equivalent, then your false positive or false negative, depending on which way you phrase it, is just going to be too large. It becomes an unacceptable situation. So, that was one comment.

The other is I think we have to realize that we're in the situation where we've got small sample sizes and with the individual bioequivalence with replicate design you're talking about further decreasing the sample sizes. So, any thought that we're going to reliably pick up interactions, unless it's just sort of luck of the draw, I think becomes a real question.

So, finally, with regard to the discussion topic 1, I would suggest that the period either be where it was suggested or we simply add, "unless there is a compelling reason not to, for an interim period of another year." But I certainly don't think we're in a position to say that "until a final decision is made to use bioequivalence." I think it would be as to whether or not to use it and, if so, in which situations.

DR. LEE: Thank you.

Dr. Bolton and then Professor Endrenyi.

DR. BOLTON: When we first made these recommendations a year ago, I personally expected to see more data than we're seeing, and we were supposed to do that so we could look at the data and decide what's going on. Well, it's pretty clear that we still don't know very much what's going on.

So, there are a couple of recommendations I would make. One is that we continue to do this the way we've done it for the next year, just to see if we can get something more, until we can make a decision one way or the other.

One thing that I was very interested in -- I know that FDA has taken that topic up with Larry taking it up on looking for mechanisms because these interactions are very fuzzy. I mean, other people have said that too. The .15 is sort of very arbitrary. It depends on sample size. It depends on the assumptions of normality. If you have lack of normality, you can induce some of these things. So, it's very hard to take them seriously unless we can find a reason why they've happened. And I know that's what the FDA is trying to do, and you did it in a couple of cases I saw in the handout.

But I'd like to see an interaction statistically and then tell me why that happened. That should not be difficult to do. If you have a strong

interaction, by looking at the formulation and knowing the physiology, one should be able to find that with some degree of reliability. I'd like to see more of that.

And I'd like to see the committee or somebody make maybe new recommendations for this next year based on what we've seen now on what to do about the things that are popping up here.

DR. HUSSAIN: The studies you saw of what Mei-Ling presented is the work we did trying to understand the mechanism of subject-by-formulation interaction.

But before I talk about that, let me share with you a formulator's perspective on this in the sense, yes, we're talking about subject-by-formulation interaction, and if you've identified something, we can correct for that.

With that in mind, we started very simple experiments. We created formulations for three components: water, drug, sucrose or water, drug, sorbitol. Two different excipients, two different attributes. We did the work at the University of Tennessee. We had an hypothesis of what might happen with respect to GI physiology. But when you go through that analysis, even with that simple formulation, it's not easy to identify what the root cause is. In fact, what we had anticipated, I think the mechanism is probably very different from that.

The point I'm trying to make here is this. You

can't get much simpler than that formulation, and if we anticipate or we expect we have a mechanistic understanding of the basis for this interaction for complex formulations, I think that's not really feasible at this time.

DR. BOLTON: To answer that, I understand the dilemma you're in, but I think this is an exercise in futility. The whole thing. Because interactions are going to pop up and we're never going to know are they real. We don't have big sample sizes. One person may have caused this. It's going to be very frustrating.

DR. LEE: Laszlo?

DR. ENDRENYI: I would like to follow up on Dr. Barr's consideration about aggregate criterion. At the Montreal meeting, several statisticians -- and they did not include me on the roster -- argued against the aggregate criterion. They suggested that even if IBE is to be studied, it could be done much better by a disaggregate procedure. But to study an issue such as subject-by-formulation interaction, IBE is not needed at all. So, I really question this aggregation of the two issues.

Secondly, I obviously do have the reservation whether subject-by-formulation interaction can be studied from these small sample sizes.

Thank you.

DR. LEE: Thank you.

Bill, you've been pretty quiet.

DR. JUSKO: I think it's pretty clear that the FDA should continue using the average bioequivalence. I have concern that the IBE criteria has a number of artifacts within it and concerns that are separate matters from the underlying science that we want to unravel. I think, as Bill Barr indicated, that there are many opportunities that we should take to try to understand reasons for variability and keep that foremost, but perhaps not throw out the baby with the bath water. More needs to be investigated in this area about variability, but perhaps this criterion has too many faults within it to be used in the manner proposed.

DR. LEE: Are you proposing to hold off throwing out the IBE?

DR. JUSKO: No. It seems like alternatives need to be investigated that allow one to characterize reasons for inter-subject variability in the context of repeated design BE studies.

DR. LEE: So, IBE is not suitable.

DR. JUSKO: That's my impression.

DR. LEE: Avi.

DR. YACOBI: I think we have heard great presentations this morning and this afternoon. I know that many of us think that IBE has definitely use and the use

has been as it has been discussed, since the early 1990s, how to do bioequivalence of highly variable drugs and highly variable drug products.

But now also we have had concern and the concern was maybe in the mid-1990s that there is subject-by-formulation interaction. Many of us thought that this is really a theoretical concern, and there were proposals to come up with data in order to prove that this real factor, subject-by-formulation interaction, is for real.

It's very nice to see new data, and my feeling is that we are hearing, even the agency, that it has a fresh look at this subject-by-formulation interaction. While it is there and we are recommending a factor of .15 or greater, but not always that should be a criteria to reject an IBE study.

My point is if subject-by-formulation interaction is not for real and we have not been able to substantiate it, then there is no really need for individual bioequivalence study. The individual bioequivalence study has been proposed from the practical standpoint in order to test the highly variable drugs not in a large number of subjects of 50, 60, 70, 80 or 100, but rather, as I recall, that well, we will do the studies in 12 subjects or 16 or 24, four-way crossover studies, two-period, two-sequence, four-way crossover study in order

to come up with data and simplify matters.

So, I hope that we are going to get to that situation where we are going to implement or we are trying to recommend that people will do IBE for highly variable drugs in a smaller group but implement the true IBE analysis because doing replicate analysis without IBE benefit, it doesn't make sense to me.

We wanted to do a highly variable drug. I wanted to do an IBE study. People came to me and said you need somewhere between 54 to about 68 subjects, four-way crossover study. So, I asked the question if I want to do just the average bioequivalence study, how many subjects do I need, and they said about 70, maybe a few more. So, I said, what's the logic of doing the replicate design analysis when I can do it for less with average bioequivalence studies? Because a replicate design study also is going to introduce additional variability in this study, and I feel it is not needed. In some of the studies here we have seen, we are seeing 50-60 subjects in the replicate design. So, from the practical standpoint, I think we have to think about it and we have to put some common sense in what we are doing and how we are going to approach this subject.

DR. LEE: Thank you.

Leon?

DR. SHARGEL: Yes. I'd like to address the first topic about whether it's reasonable to use average bioequivalence. I certainly agree with most of Les Benet's comments.

One thing. Generics have been on the market for over 20 years using average bioequivalence since Waxman-Hatch in 1984 more formalized the approach of ANDAs.

Being in the academic, as well as in the generic arena, I am very much aware that our innovators have looked at differences among the generic and the branded. They have not published and they have not pushed it out that widely because they haven't found as much, and they spent a lot of energy with the products coming off the market right now. Obviously, they're looking at a lot of differences between formulation effects, drug substance effects, clinical effects, and everything else. And we've had these arguments with NTIs as well.

We've also had the arguments going back 20 years, and one thing about being older is that we did originally use normal, healthy males, usually nonsmokers. We were worried about enzyme induction and things of this sort. So, many of these older products were based on the fact that we were really looking at differences between drug performance in terms of bioavailability between the two products, not so much as clinically. The argument was

it would be more appropriate such as a highly variable drug.

DR. LEE: Let me take three more questions. Then I would like to sum up what I heard. I think Marv Meyer had his hand up, and then Sandy and Laszlo.

DR. MEYER: This is quick. It seems to me maybe we have a nomenclature problem with it that's raised expectations. We talk about individual bioequivalence and we talk about subject-by-formulation interaction, and I didn't hear a single presentation that really identified subject X or individual Y and said this really means for sure that there's an interaction or I know anything about them. I think we would like to think that we're going to somehow identify that my grandmother is going to be different than your 12-year-old son in these studies, but it ain't going to happen. Until we figure out a way to utilize IBE better or to study that phenomenon better, it's not going to be very useful.

DR. LEE: Sandy?

DR. BOLTON: I just have a comment about sample size. Number one, some of the reasons why one passed and the other failed, using IBE and average, might be just a function of sample size. Number one.

The other thing is sample size for variable drugs -- I want to expand what Avi said. I agree with him

100 percent. First of all, you're limited to very, very highly variable drugs, which is a small subset of drugs, and even then I am not sure that we do better on individual bioequivalence. I wish somebody would look into that a little further to see if we really have an advantage with variable drugs using individual bioequivalence and where that cutoff point is. Once we were told it was 30 percent is an advantage. Then it was changed to 45 or 50. My sense is it's even bigger than that.

Finally, I'd like to say one thing about Les' final comment about reducing the limits for a public relation point of view. I'm against that because I think that the generic industry, if they made an effort, could make a campaign to explain in lay words to the doctors and the public that, indeed, these generics are not 50 percent different than the brand name, which many doctors think they are. So, that could be done without having to change the limits.

DR. LEE: Thank you.

Laszlo?

DR. ENDRENYI: Just to clarify on this point to Avi, Leon, and now to Sandy, for highly variable drugs that is necessary and does the job is scaling, reference scaling. It's not individual bioequivalence. It's scaling. And scaled average bioequivalence does a much

better job at that. So, I don't see the role in this, for highly variable drugs, of individual bioequivalence.

DR. LEE: Well, it seems to me that there's a consensus to continue using the ABE.

Larry, you would like to make a comment?

DR. LESKO: Yes, I'd like to make a comment.

DR. LEE: Just very briefly.

DR. LESKO: Just briefly comment? All right. That will be harder.

DR. LEE: One minute.

DR. LESKO: I wanted to talk about the current situation, and the current situation as the agency has to make a decision when given an application to review.

We have in our current guidance that sponsors have the option to explain why they would use another criterion other an average bioequivalence. The most logical extension of that is the sponsor that requests to use IBE for a highly variable drug.

We've heard today and some of the data we presented was that the aggregate criteria under IBE gets you to a win under that scenario with many different combinations of numbers representing the means differences, the variance differences, and the subject-by-formulation term. And you can mix those all up and come up with a win with different combinations.

Some of the combinations create concern in our mind where we give a tradeoff on the mean difference with an increase in variability to test and maybe even a subject-by-formulation interaction, and it says pass. That doesn't seem acceptable. So, some of the combinations of numbers don't seem to make sense intuitively to prove a product using IBE.

So, under the current situation, if the sponsor were to come in without any constraints getting to this discussion topic number 2, we would then have a situation where we can approve a product that may differ from a reference product having up to 125 percent of the bioavailability or as little as 80 percent of the bioavailability if there's an appropriate reduction in the variance of the test product.

We can also have a product which we might reject that would have 90 percent of the bioavailability, but we would reject it because the within-subject variance for the test product is a little bit higher than the reference.

So, it gets kind of confusing. But the point is, without constraints, I'm concerned that we'll be in a position to make a decision on a product that has a different bioavailability than the reference and may even be exhibiting a subject-by-formulation interaction when

they scale it and it'll pass. That's why we put in the constraints.

And there's something illogical about that. We created a method where we all agreed, at least in 1999, to look for subject-by-formulation interactions. Now we have a method that's picking them up, and we're saying let's pass the product.

So, I think we need a constraint under the current situation if we're going to implement the individual bioequivalence in our current guidance. If we're going to retain the scaling benefits of that equation, which we can do with the constraint on sigma D, it will make it a bit harder, but you can still retain the scaling benefits of the equation. Then I think it makes sense to put that constraint in there.

I think also we want to bring the differences in the test-to-mean ratios down to 15 percent, and there is a sort of quasi scientific reason to do that. It's to pretty much bring the differences in mean under the average bioequivalence scenario that we would approve under average in line with the IBE so that at least in the short term, we don't make any decisions we might regret in the long term when we have more data and make a final decision on using IBE for the marketplace.

So, I think that's why the constraints are

important, and because when we leave today we have to make a decision on that guidance in the face of these replicate design studies, I think we have to come to some resolution of that because if you say don't put any constraints on there, then we're going to be faced with a difficult decision of making that decision for the marketplace.

Now, if you think a bit further, if we let this occur with generic product number 1, using scaling and using these larger mean differences to occur, what do we say about two generics in the marketplace? Are they going to be more inequivalent than they might possibly be under an average bioequivalence scenario? Well, I don't think we want that. But this criteria without constraints I think will create that probability that two products on the generic side could be more different than they might be under the average bioequivalence scenario.

DR. LEE: Discussion topic number 2.

DR. LAMBORN: Could I ask a clarification question?

DR. LEE: Yes.

DR. LAMBORN: The comment was made that we can do scaling using average bioequivalence. In today's environment with the existing guidance, is there a scaleability criteria outside of the individual bioequivalence situation?

DR. LESKO: I don't believe we've explored that. We'd have to explore that as a possibility.

DR. LAMBORN: So that the statement that that could be done, it is not currently part of the guidance.

DR. LESKO: It is not currently part of the guidance, and our working group has not spent a lot of time looking at that.

DR. LEE: So, I was going to say that there is consensus that we continue to use the ABE. What I heard around the table is that there is a lack of consensus about what IBE is all about. Dr. Moye suggested there are other ways to look for that, and there's some suggestion we should throw out it entirely. There's some sentiment that maybe it's premature.

Are we ready to propose to consider the option until we understand under what conditions would IBE be appropriate for market access?

DR. BARR: Excuse me. Are you asking whether or not we think that IBE ought to be allowed as an alternative criteria or whether or not it ought to be allowed to be continued to be studied? What is the question, Vince, that you're asking us?

DR. LAMBORN: Is this question 2, discussion item 2 that you're on now?

DR. LEE: No. Question number 1 is that we

need to come to some decision, provide some advice to the agency about how they should proceed. The proposal is should ABE continue to be used for another year until a final decision is made to use IBE for market access.

DR. LAMBORN: Could I suggest that in order to answer question 1, perhaps we need to discuss discussion item 2 because I think that what's being expressed is a concern about -- there's an implication in 1 that we would continue to allow IBE to be used for the exceptions. And I think to say that we would continue to study with replicate designs, implying that they could use the IBE, I think we need to address Larry's concerns that he just raised. So, I would propose that need to address discussion item 2 and then come back to the vote.

DR. LEE: Okay.

DR. KIBBE: Larry, just getting back to the concerns you raised, I only spoke to item 1, and my issue basically is I think one year from now I'm not going to be comfortable converting everything over to IBE.

DR. LESKO: We're not proposing that. No. We haven't proposed that we're going to convert everything to IBE. We recommend replicate design for two classes of drug products.

DR. KIBBE: The statement, if I read it, reads that we will do it for another year until a final decision

is made to use IBE for market access. My point is the statement ought to read that we're going to use ABE for market access unless there's a compelling reason to use a different system.

And then my question to you is -- and I'm following up on what Kathleen has said about topic 2 -- are the criteria that's currently listed in topic 2 good enough, or do we need better ones, in your opinion, than that in order to make IBE a viable alternative to ABE?

DR. LESKO: To clarify the first point, because I think it's important, the context for discussion topic number 1, is the current guidance in which we recommend replicate design for two classes of drug products, modified-release, and highly variable. I think in 1999 and then when we subsequently put out the guidance, we made the decision that we would not recommend replicate design for the other classes of drug products, and hence IBE would not be the way to market access. So, that's discussion topic number 1.

On discussion topic number 2, we think those constraints on the IBE criterion would make us comfortable to approve a product on IBE, which would include a measure of scaling, but it would exclude approving a product that deviated in its mean ratio to a degree greater than we currently allow under average bioequivalence. It would

also signal to us that if we had a high value for sigma D, which could indicate a true subject-by-formulation interaction or perhaps a group-by-formulation interaction, that would not be then an IBE criterion for market access. One would go back and use average bioequivalence if it passed under the criteria.

DR. LEE: Well, I guess the discomfort is that there's a perception on Art's part that the IBE would eventually be replacing ABE.

DR. LESKO: It's looked at as an alternative for a sponsor to make a choice a priori whether they want to use average and IBE. We don't envision it as a replacement for average bioequivalence, at least not at the present time. In each case, whether one picks the average or the IBE, there's going to be both a producer of risk of success and failure and another risk of success and failure in terms of a patient risk.

DR. KIBBE: Your concern about criteria was that you thought you heard us saying that we were going to change the criteria as listed in 2?

DR. LESKO: No. That wasn't my concern. The criteria listed in number 2 is what the working group is recommending for consideration as the prerequisites to utilize IBE for market access.

DR. LAMBORN: To clarify, I think some of the

items in 2 would be changes from the existing guidance. Is that correct?

DR. LESKO: That's true. The existing guidance, for example --

DR. LAMBORN: So, the issue is are we prepared to support the proposed changes in the existing guidance.

DR. LESKO: That's correct. The main changes on the test-to-reference ratio is constrained to 15 percent rather than 20 percent. The current guidance does not have any constraint on the value of sigma D, subject-by-formulation interaction. All of the other things on there, the other four bullets, if you will, are in the current guidance. That's nothing new. So, there are two new bullets on there compared to the current guidance.

DR. LEE: Laszlo, are you going to help us out of this dilemma?

DR. ENDRENYI: On discussion topic 1, if it would state, as already suggested, that is it reasonable and appropriate for FDA to use average bioequivalence for market access unless there is a compelling reason not to, period, end, I think that would still permit the investigation of IBE under discussion topics 2, 3, and 4.

DR. LESKO: That's logical to me. It's removing a time frame.

DR. LEE: Is the committee comfortable with

that?

DR. KIBBE: Yes.

DR. LEE: So, we just put a period where?

DR. KIBBE: After "to."

DR. LEE: "A compelling reason not to."

DR. KIBBE: Period.

DR. LEE: And then period. That would still allow us to go and discuss item number 2.

Discussion topic number 2. Yes, Laszlo?

DR. BOLTON: I just have to a question to clarify. Are you saying that you have an option here? If it doesn't pass these, you can use average bioequivalence. If that passes, then you're stuck with this.

DR. LESKO: No. We're not saying do the study and play a winner.

DR. BOLTON: Yes. If you choose this, you must pass.

DR. LESKO: The guidance is very specific in saying that the sponsor should choose a priori in their study protocol which methodology they're going to use.

DR. BOLTON: And they will use these criteria as new criteria.

DR. LESKO: That's correct.

DR. BOLTON: Okay.

DR. LEE: Now we're on discussion topic number

2 on the criteria.

DR. ENDRENYI: Could I take a rain check because the gentlemen handling the slides just went out?

DR. LEE: All right.

Please.

DR. ZARIFFA: I'm looking at discussion topic number 2, and I'm framing it in my mind as how do we collect more replicate design data sets while disallowing concerning patterns under IBE? So, there are two points that follow from that. The first is, how much more will we gain from an additional 10, 20, 50, X number of replicate design data sets? And two, do the additional constraints that we're putting on to disallow concerning patterns actually make sense?

So, there are two pieces that follow on from the question. The first piece has to do with what is the value of the additional data, and Marv asked this earlier. Don't we know enough? Don't we have enough? Haven't we simulated enough? And that comes to discussion topic 4. So, I'll leave that to one side.

The question of whether or not the additional constraints make sense in the short term -- we're talking about possibly just a year -- is something that we should keep in mind. Personally I was swayed by the arguments that Laszlo and Kam put forward regarding the geometric

mean ratio, and I would hate to see this community take several decades back in time by going to essentially what comes down to look at means in small data sets. I don't like that.

And the question about the constraint on sigma D being .15, it's been demonstrated over and over again that that is not valid under a number of different assumptions which arise quite naturally in practice.

So, those would be the two points, and the rest I'll table for discussion topic 4.

DR. LEE: Okay. Let me take the chair's prerogative and put the microphone back to Marvin Meyer for us to hear his opinion.

DR. MEYER: From what I understand, we have something like six bullets under topic number 2. I don't think there's any debate on it should pass IBE criterion for a study done under IBE, although I'm not real clear what criterion we're going to use, but whatever that must be, then we will use it.

24 subjects is fine.

I think there's debate whether there should be no significant subject-by-formulation interaction. That shouldn't be a reason to dump a study, I wouldn't think, if it's above .15. Rabi showed some data that suggested that didn't mean a heck of a lot.

A constraint? Personally I believe we ought to have one. Laszlo, I think it was, presented some data. Les recommended I think a 20 percent for Cmax. Some constraint. Now, whether it's 15 percent, it's 20 percent, I don't think we want to go above 20 percent, and maybe not above 15 percent because I think the perceived differences -- now, we're going to have to set ourselves back perhaps, but at the same time, we don't have to worry that the agency has approved some products that shouldn't have been approved because we have too lax of an approval process. I think we can expand that, make it up to 20, 25 percent, if necessary.

I object a bit to the heterogeneous population. If you think about it, what does that really mean? That means blacks/whites, males/females, old/young. That's eight permutations. With 24 subjects you could have 3 of each of those subgroups, and I don't know what that will tell you. So, I don't know we're going to achieve that objective. I wouldn't think we should allow all young, healthy males. We should have a little more diverse population, but to mandate some prescribed heterogeneous population I don't think will work.

So, those are my comments.

DR. LEE: Laszlo, you want make a comment?

DR. ENDRENYI: First of all, I would like to

repeat Kam's plea. That was the most important one. Do not introduce a new regulation until you've studied fully the science, please. So, thinking of new criteria before they have been studied I think would be deadly, disastrous.

Slide 9.

DR. LEE: And this slide would address topic number 2?

DR. ENDRENYI: Yes.

DR. LEE: Okay.

DR. ENDRENYI: As already indicated, I'm very strongly against the 85-117 percent limitations. As Kam says, that takes us back. Furthermore, I believe, as far as I can make out without additional studies, it will be not an individual bioequivalence criterion, but a GMR criterion like Canada for Cmax.

We haven't talked about modified-release formulations. The sigma D criterion. The .15 is not appropriate. It's true that in the model it is sigma D and sigma W -- that's the within-subject variation -- are independent. When they are estimated, estimated interaction, estimated variance are not independent. They are directly related, linearly related in fact. So, a simple set criterion is not appropriate. It will do absolute injustice to highly variable drugs.

Furthermore, there are some other problems like

sensitivity and what we already talked about, being able to have the sensitivity to be able to detect an interaction in small groups. It has a problem, but there is a basic problem with the sigma D for .15.

We haven't talked about modified-release formulations, and I think there are some basic points here. The modified-release has subgroups. Delayed-release with lag time, usual kinetics; extended-release, usual kinetics, slow absorption. For these, there is no reason whatsoever to require replicate design studies. For sustained-release, controlled-release, there may be for investigational purposes. But why?

DR. LEE: I think we got your point.

DR. ENDRENYI: Actually there was one other point.

Replicate design. And I think we talked about individual bioequivalence, but there is also a point about the replicate design study. Why do we want it? My sense is that we want it apparently for the sake of data collection. Question: For regulatory purposes, is this a need to know for regulatory approval or is it nice to know to get data? It would be useful to clarify this point.

Thank you.

DR. LEE: Thank you.

May I have the committee express the opinion

first?

DR. BENET: Vince, can I make a comment?

DR. LEE: Yes, Les.

DR. BENET: I want to come back to both what Nevine said and what Laszlo said about the GMN and Kam's position on the point estimate, the GMR. Basically we are not asking for any new criteria. This is not an untested criterion. Nightingale and Morrison in 1987 looked at 224 products, one of which was out of plus or minus 15 percent. Gene Haney summarized a couple of years ago since then what the rule of change -- none of them were out of plus or minus 15 percent. So, we're not adding any new criteria because the present criteria have always maintained it within that area.

Why I want plus or minus 15 percent on the IBE is because exactly in opposition to what Nevine and Laszlo and Kam said, this is new. We are doing something new with IBE. We are not doing something old. So, it is not that we're doing something that was different than the past; it is that we have a new way that we're going to approve drugs.

And I think it's important, as a number of other people have said, to make sure that the clinical community and the patient community -- I know Nevine, as a statistician, says that's not important, but I can tell you

it is important and it's important for the people in the United States that they believe this.

And Sandy, you're crazy if you think that the generics can get these clinicians and make them believe because the generics don't put the money into the pocket of the clinicians. So, you've got to deal with reality.

And I do not believe that this is something new. I believe it's exactly what we've been doing in the past.

Thank you.

DR. LEE: Thank you, Les.

I think that we do need to move along, and I would like to ask the committee to express their opinion about topic number 2. Of all the criteria proposed, which one might need some more discussion?

DR. LESKO: Vince, could I clarify something?

DR. LEE: Sure.

DR. LESKO: I'll give it to my colleague.

DR. HUSSAIN: Well, I think the constraint on the data Professor Benet talked about was essentially historical data that we have looked at. Mean differences between approved generic products and so forth are very tight. I think that's what he was referring to.

DR. BOLTON: Can I just say one quick thing? If you start adding these restrictions -- I'm against

adding those restrictions -- then the whole properties of this metric are changed. So, now we have to reevaluate what that metric really means with these new conditions. I don't think it's fair to just arbitrarily do it, to just throw it on there and say, well, that's good. You're making up numbers. That metric came from scientific basis, whether we like it or not, and now we're making it a completely different thing. It's not the same anymore. So, why not come up with a different criterion that makes more sense to everybody?

DR. LESKO: Vince, I think it's important to clarify one other thing, if I could, on this debate about the equation. I've heard it a couple of times, but I still don't understand we're going back 10 years.

But that aside, if you think about the IBE equation, what we're saying is we're putting a constraint on what's in the parentheses comparing the mean of the test to the mean of the reference. We're not changing the right-hand side of the equation. The right-hand side of the equation stays as it is, natural log of 1.25 or whatever.

Laszlo made the comment in his presentation, by putting a constraint on that, that's going to eliminate some of the width that would be allowable for scaling, but you did say you're not sure whether it would be a GMR or

would it be a true scaling.

Now, if we converted that to a linear scale and we have to do that, I don't know what the tradeoff would be by putting in a constraint on that one part of the IBE. I mean, we do it now. We have a 20 percent constraint in our guidance on that parentheses, and what we're saying is let's make that difference in the parentheses 15 percent, not changing another part of the equation. It may change the properties of the equation. We can explore that, but I don't think it changes them significantly.

DR. BOLTON: [Off microphone] how that changes. Do a little study and then say, listen, doing this doesn't change things very much. It might be more appealing, but we don't know that.

DR. LEE: May I consult with the statisticians on the committee? Yes, Kathleen?

DR. LAMBORN: What is the statistical question you were going to ask? I was going to comment on something a little different.

DR. LEE: Whether or not this is statistically sound.

DR. LAMBORN: I'd like to sort of split this thing into two parts. I think if the statement is in moving from average bioequivalence to individual bioequivalence, we don't want to allow to pass products

that are further from a ratio of 1 than we had before, then I would say that's just a comfort level with regard to what we're doing. Clearly by adding an extra constraint, it will reduce the likelihood that something is going to pass. From the sounds of things, it shouldn't make a case where something that would have passed under the old rules would not pass now because under the average bioavailability, they're passing anyhow.

I guess the thing that I'm coming down to is the agency is seeing, now that they've had a year of experience with the individual bioavailability, that they're not comfortable with the guidance as it stands. It's almost like we've got a choice. We either say withdraw the option of using IBE until it's been studied more fully, or put some constraints on it so that there's a comfort level until you've had a chance to do the additional study to see what the impact is.

But I think clearly if the people who are seeing the data coming through are not comfortable with what they're seeing and feel that it could potentially be allowing something unsafe through, something has to change. So, that's partly a statistical answer and partly just my own personal opinion.

DR. LEE: Does your colleague next to you have a comment?

DR. MOYE: If I understand the statistical question, I would say that this new methodology is unsound for identifying what it has claimed to identify, that is to say, for identifying a subgroup-formulation interaction.

I would say that it is sound methodology to identify something that so far, to my knowledge, hasn't been detectable, and that is this notion of a subject-by-formulation interaction. So, if we're looking for demographic and subgroup interactions, then I think this methodology should not be used.

What it has been specifically designed to evaluate is an effect that I understand has not yet been identified and that is this ephemeral subject-by-formulation interaction that is exclusive of, separate and apart from ethnic or gender formulation interaction.

DR. LESKO: I'd like to respond to what Kathleen said. Without trying to rephrase it, I think she put it in perspective. It's exactly what we're worried about and it's exactly why we want to put the constraint as we've suggested it.

I just did a quick look also down the table of data that was new that we presented to the committee under the ABE column, which shows the ratio of test to reference means. They're all very tight. We're not even close to 15 percent on any of them. So, there is a lot of worry about